The Science of the Self

A Summary of Current Literature

by Michael W. Taft

A

being human™

Book

The Science of the Self

A Summary of Current Literature

Table of Contents

Introduction——**Page 1**

The Current Model of the Self——**Page 3**

The Embodied / Emotional Self——**Page 6**

The Mnemonic / Fictional Self——**Page 13**

The Developmental / Evolutionary Self——**Page 21**

The Social / Cultural Self——**Page 24**

Conclusion——**Page 32**

Appendix——**Page 33**

Notes——**Page 34**

The Science of the Self

A Summary of Current Literature

by Michael W. Taft

Funded and distributed by The Baumann Foundation
1770 Post Street #185
San Francisco, CA 94115

info@thebaumannfoundation.org

A

being human™

Book

Introduction

"Knowing others is wisdom. Knowing the self is enlightenment."

– Lao Tzu

Perhaps there is nothing so elusive in human experience as the self. Everyone experiences the self as the center of their life and yet cannot seem to find it. But it is not for want of trying. Humans have long been fascinated with the self and have made tremendous efforts, usually either philosophical or religious in nature, towards understanding it,

In many traditions the self, or essence of a person, is called the "soul." Socrates and Plato (5th century BCE) believed that the soul was an immaterial occupant of the human body, which departed at death – a belief common in many cultures. In his book *De Anima* (On the Soul) Aristotle wrote that the soul represented a human being's capacity for rational activity, "like an axe has an edge." He felt that the soul resided in the physical heart.

Continuing from Aristotle, the Persian philosopher Ibn Sīnā[1] (11th century CE) defined the soul as "what a human indicates by saying 'I'." [2] While languishing in prison, he devised his famous "floating man" *Gedankenexperiment*, in which he asked readers to imagine that they have just been created "at a stroke," floating in space, unable to see or feel anything. Under these conditions, Ibn Sīnā argued, there is no doubt that a person "would affirm his own existence, although not affirming the reality of his limbs or inner organs, his bowels, or heart or brain, or any external thing."[3] In other words, the essence of a human being is consciousness or self-awareness – an argument that would eventually became very powerful in Europe.

Aristotle's and Ibn Sīnā's ideas were picked up and refined by Thomas Aquinas (13th century), who insisted that the soul was absolutely immaterial. It is this notion, then, of the soul as a non-physical entity, the ghost-like essence of a human being, the center of consciousness and self-awareness, that was brought into the scientific era by René Descartes (17th century) in his famous dictum, *cogito ergo sum*. Descartes takes thought (i.e. consciousness) as equal to identity. It is because of Descartes that science for so long felt that the mind was separate from the body and the brain, often called the "mind body problem."

With the rise of psychology in the 19th century, these philosophical ideas about the soul began to be looked at with the eye of science. Psychology, which literally means the "science of the soul," dropped the word "soul" from its vocabulary, due to its religious connotations, and instead began to use the terms "consciousness," "self-awareness," or simply, the "self." Psychology as a science began with Wilhelm Wundt, who conducted experiments to understand how his mind worked. He and his followers, the "introspectionists" sought to investigate conscious experience and analyze it into irreducible components. Their method (as the name suggests) was to examine the subjective experiences of themselves or their patients in an attempt to determine how the mind works.

By the early 20th century, the introspectionist's methods came under attack. Because they were simply reporting on subjective experience, their results were unverifiable and therefore unscientific. The result was a new, scientifically valid direction in psychology, one which looked solely at externally observable events, or behaviors. This movement, known as behaviorism, dominated psychology for several decades. Behaviorists believed that consciousness couldn't be studied because it couldn't be observed. The most extreme among them went so far as to insist that consciousness and the mind didn't exist at all! A human being was a machine, a body and nothing more, with mental states – if they existed at all – being nothing more than a side-effect of its mechanical operation. A stimulus is applied and a response is seen; this is the behaviorial view of a human being.

But between the application of the stimulus and the occurrence of the response isn't it possible that something is happening in the brain? With the rise of computer science in the 1950s, some psychologists felt that mental operations could be compared to the computations done by computers, and that these cognitive operations could be legitimately determined. "Cognitivism," as this new school of psychology is called, still rejected introspection (including the work of Freud and Jung) as scientifically invalid, but felt that the mental states did indeed exist, in the form of cognitive operations. In this view the mind was seen as an information processing system, and could be investigated by studying its cognitive operations, such as how sensory data are input, stored, transformed, and elaborated in the brain.

The cognitive approach revolutionized psychology, and has been very influential in many other areas, such as linguistics, artificial intelligence, and – importantly – neuroscience. Cognitive neuroscientists have used the information processing concepts of cognitivism to make great strides in understanding how perception, attention, memory and thinking relate to underlying brain mechanisms.[4]

While the cognitive movement has resurrected the existence of internal, mental states, however, it has not been successful in creating a science of the whole mind. First, cognitivism concerns only the cognitive (i.e. thinking) parts of the brain, and does not look at emotions or motivation. A mind may be something like a computer, but this analogy falls apart when we attempt to include affect and goal-orientation. Furthermore, cognitive neuroscience has been able to demonstrate how memory, perception, and so on operate, but not in how they function together. Finally, the cognitive approach is general and cannot account for how genetics and life experiences shape the *unique individual* workings of a brain.[5]

Thus in the last decade a new approach has arisen which seeks to scientifically understand the biological basis of the mind in its entirety, including the self. This approach has yet to acquire a name, but it does have several characteristics in its view of the self which are new, and which may well be defining of its viewpoint. For the sake of this paper, I will call this viewpoint the "Current Model of the Self" (CMS), which is a name of convenience only.

The Current Model of the Self

This paper will briefly summarize what I believe to be the major themes of the Current Model of the Self. Then it will review the ideas of the foremost scientists in the field, and show how they contribute to the CMS.

As noted earlier, the CMS has arisen out of the previous cognitivistic work. One important way in which the CMS does not differ from this earlier model (or behaviorism, for that matter) is the notion that the *self arises wholly as an organic function of the brain*. There are no spiritual or metaphysical aspects of the self. It can be explained in its entirety by understanding the operation of its physical components.

The new aspects of the Current Model of the Self can be summarized in five basic concepts:

1. *The self is distributed*.

In naïve human experience, the conscious self seems to be a single, unified entity. Unless there is a major psychological pathology, an adult human feels his- or herself to be *one person*, not many, and this conviction is central to his or her personality. It is not surprising, then, that for thousands of years the majority of theological and philosophical speculation has regarded the self as a singular entity, the very meaning of individuality and singularity.[6] Descartes speculated that the self must be located in the pineal gland because this gland is a singular structure centrally located among the many duplicate structures in the brain.[7] Even as late as the 1970s, Carl Roger's saw the self as a unity (albeit a "gestalt").[8]

Virtually all the major researchers (including LeDoux, Gazzaniga, Damasio, Dennett, Wegner, and many others) now agree it is very unlikely that the self is located in a single brain system. Most data suggest that the self arises from the combined output of many different modules located throughout the brain. For some researchers (Gazzaniga, 2003) the self exists as a network of associations, whereas Damasio believes that some of the modules refer to body awareness and emotional tone. Perhaps the ultimate limit of this conceptual direction is reached in Joseph LeDoux's theory of the "synaptic self," which states that the self arises ubiquitously in the synapses of the brain, virtually everywhere.[9]

It appears that we are not going to find a single "self spot" in the brain. The self is composed of a large number of relatively independent modules, each of which evolved to solve unique challenges.

The "disappearance" of the self into a mass of competing brain modules is probably the most intuitively unsettling notion of the CMS. In 1996, as this concept was beginning to come clear, the evolutionary biologist William Hamilton remarked:

> "In life, what was it I really wanted? My own conscious and seemingly indivisible self was turning out far from what I had imagined and I need not be so ashamed of my self pity! I was an ambassador ordered abroad by some fragile coalition, a bearer of conflicting orders, from the uneasy masters of a divided empire….As I

> write these words, even so as to be able to write them, I am pretending to a unity that, deep inside myself, I now know does not exist."[10]

2. The self is largely implicit.

The CMS recognizes that some aspects of the self are unconscious. Because the term "unconscious" has such a long, baggage-ridden history of use in psychology, neurology (particularly memory studies) has opted instead to use the term "implicit" when referring to information not available to consciousness, and "explicit" for its opposite. This paper will use both sets of terms interchangeably.

The idea that the self is composed of both explicit and implicit material is of course not a new one. Some ancient philosophical models of the self (i.e. Buddhism and other Asian traditions) were clear on this point, and modern Western psychodynamic theories (Freud, Jung, and so forth) are predicated upon the existence of the unconscious. As mentioned earlier, the behaviorist school doubted the existence of even the *conscious* self, let alone the unconscious – and this position was ultimately found to be too extreme. With the rise of the cognitive school, the explicit self once again became the subject of serious study. But it has only been with the growth of the CMS that the implicit self has come under the microscope.

By the mid-1990s, several major researchers (Gazzaniga, 1998; Damasio, 1999) published major works about the implicit self. These pioneers made extensive study of brain-damaged patients which clearly demonstrated that a great deal of information about the self is not available for explicit inspection. It is functioning behind the scene, so to speak. Their work further shows that it is probable that the *great majority* of the self is implicit. "98% of what the brain does is outside of conscious awareness,"[11] says Michael Gazzaniga.

A simple example of the implicit workings of the self can be found in studies of egocentric bias. It is a well-established fact that people will distort past events to make themselves look better, and yet they are *consciously unaware* that any distortion has occurred. There is some implicit "machinery" of the self at work here, secretly doctoring the records to shore up the confidence of the explicit self. Tests of amnesiacs with severely damaged conscious selves reveal that these implicit bias mechanisms remain unchanged.[12] Thus the implicit self happily chugs along, even if the explicit self is gone. We will look at many more examples of the implicit self in the course of this paper.

3. *The self is very fluid and to some extent fictionalized*.

As the example above shows, one interesting result of having a self that is divided into implicit and explicit portions is that we may not be consciously aware of exactly who we are at any moment. Although humans intuitively feel their self to be stable and constant, there is a large, robust body of research data that demonstrates how the brain uses

distortions and half-truths, carried out automatically and beneath conscious awareness, to bolster the seeming-solidity, unity, and worth of the self.[13]

This fictionalization begins with the very concept that we are one person, one self, which according to the CMS is nothing more than a convenient story, when in fact we are composed of a large number of competing brain modules. It continues down to the details of our personal history, which are changed to suit the circumstances (again, implicitly), and even in our behavior. The math test scores of Asian American women change depending on whether they have been unconsciously reminded that they are Asian or they are women. The beliefs that the implicit self holds can change the actual behavior of the explicit self, all without the conscious knowledge of the self.[14] Thus who we are is changing constantly, and is often fictionalized.

4. The self arose slowly in evolution.

In ancient shamanic cultures, animals were "people" too inside, having a self or soul like a human being. European Christianity eliminated this view, depicting animals as having no souls, and therefore no selves. Descartes continued in this vein, writing that animals were nothing more than complex machines, without souls, thinking, or reason.

This bias gave the science of the self a long hangover, as it were. Up until very recently, the self was thought to arise as a function of language – something that could only take place in human beings.[15] The CMS notion that the self is composed of many modules, however, has given animals back their soul, so to speak, albeit one that is different from that of a human. The notion that the self is distributed in many nodes throughout the brain has naturally led to the idea that each of these nodes arose separately in evolution and that they developed at different times.

If this is the case, then it is logical that animals possess certain elements of a self. The nearer to a human being, the more "complete" the animal's complement of self modules. There is a substantial amount of research to support this conclusion. One example is the ability of some higher animals to pass the "mark test," which indicates that they have a sense of self.

The concept of the evolutionary self is a powerful one, because it gives science the ability to look at the fundamental workings of the implicit self in animals. It is reasonable to assume that these rudimentary aspects of the self function more or less the same in humans. Thus we can construct a model of the self from the foundations up, so to speak, starting with its most basic modules in simple organisms.

5. The self is partially embodied and emotional

Historically, the idea of the self has not only been the exclusive domain of human beings, but also strongly associated with the mind as opposed to the body. In Christian theology, the soul was completely spiritual, nonphysical, untouched by the gross earthly matter of

the body. Descartes declared his immaterial self to be a *res cogitans* (a thinking thing), and ascribed to it all the higher mental faculties. It seems that there has been a kind of revulsion at the idea that the self – the supposed essence of a human being – might somehow belong to the messy, dirty, inconvenient, painful, and physically real earth.

Yet if we are to consider that the self arose slowly in evolution then it is logical to assume that its oldest, deepest modules in the brain are not related to the higher functions. The self, such as it is, of a lizard is obviously not concerned with language and higher reasoning. Rather, the self of simple organisms must be related to things such as maintaining bodily integrity, coordinating bodily movement, and responding to emotional stimuli. These relatively simple aspects of the self would have appeared early in evolution, and yet provided the necessary push for the subsequent edifice of the self to arise.

Although not all scientists refer explicitly to this concept, the general opinion seems to be moving to include this view. Damasio in particular focuses on the role of body awareness in self-identity, and shows how the neurology of body awareness and that of emotions are deeply intertwined.[16] Several researchers speculate (Llinás, for example) that the origin of self awareness is the need to coordinate bodily movement and predict its outcome.[17] And LeDoux (2003) gives a special role to emotion in orchestrating the cohesion of the self system.

About This Paper

These five themes, then, comprise what I call the Current Model of the Self. The remainder of this paper will explore these ideas as they appear in the books and papers of major researchers in the field. One oddity is that since these scientists each work in their own domain – neurology, psychology, sociology, etc. – the information is organized along somewhat different lines than the five aspects of the CMS. Yet these five aspects appear over and over in the subsequent pages, running like threads throughout. [18]

While the CMS seems to solve many issues about the nature of the self, it brings up several new questions of its own. Probably the biggest of these is how this seemingly fragmented system of independent agents gives rise to the single, cohesive, stable sense of self. We will see that there are several very intriguing possible answers to this question.

Another big question is how the CMS affects the moral and philosophical questions of self. This turns out to be such a complex and vexing question that I will not address it in this paper, as it could easily triple it in size.

The Embodied / Emotional Self

Current neuroscience is only just beginning to come around to investigating the idea of a self that evolution has built from the "bottom up," beginning with brain modules associated with regulation and movement of the body. While this viewpoint of the self arising from body awareness has not been in the majority until now, it nevertheless has a venerable history, within which Damasio places such thinkers as Spinoza, James, Nietzsche, Husserl, Merleau-Ponty, Charles Sherrington and others.[19]

Damasio's view of the self turns the traditional neurological view on its head. Whereas the old model saw the self as beginning with language in the "higher" (i.e. more recent in evolution and more uniquely human) aspects of brain function, Damasio sees the self as rooted in the "lower" (i.e. present even in simple creatures) neurological structures, and then building up in ever more complex, subtle, and less primary layers from there.[20] In a way, Damasio's self takes some systems of brain function that have been around for a very long time indeed, and cleverly repurposes them to create something unique in the animal kingdom. As he shows, this reworking of already existing material is an efficient way to strongly enhance the evolutionary advantages of emotions and awareness itself.

Damasio's model divides the self into three basic components: the *proto-self*, the *core self*, and the *autobiographical self*.

The Proto Self

The proto-self is not a part of the conscious self at all, but rather comprises the neurological basis from which the conscious self arises. Damasio's briefest description of the proto-self describes it as, "… a coherent collection of neural patterns which map, moment by moment, the state of the physical structure of the organism in its many dimensions."[21] That is, the proto-self is the brain's representation of the state of the body. His phrase "in its many dimensions" refers not only to the many variables of body condition being tracked, but also the many and various senses the body has evolved to do the tracking.

The proto-self's mapping of the body includes input from the external senses (vision, smell, hearing, taste, and touch) as well as the interoceptive senses that allow humans (and other mammals) to monitor the internal condition of the body. The interoceptive senses include proprioception (knowing the body's morphology in space), the vestibular sense (balance and spatial orientation), the visceral sense (how your "guts feel"), and the sense of the internal milieu, including pain and temperature.

While the external senses use the fast A alpha and beta nerve fibers, which are suited for the relatively rapid and abrupt changes possible in the environment, the interoceptive senses use an entirely different and dedicated system for signaling the internal state of the body, the C and A delta fibers. The C and A delta fibers are ubiquitous throughout the body and represent a much older and slower system, better suited to the relatively stable and nuanced internal environment. These nerves respond to the amorphous chemical

washes that continuously bath internal tissue and are sensitive to changes in such things as pH, partial pressure of O_2 and CO_2, glucose, lactic acid, glutamate, histamine, and serotonin. They also respond to temperature, pressure, the flush of the skin, itches, tickles, sensuous touch, and genital arousal.

Beyond the symphony of nuanced feedback flowing through the C and A delta fiber system, there is also the *chemical* sensing of the internal state of the body, which flows through the bloodstream into the special receptor sites in the basal ganglia that are not screened behind the blood-brain barrier.

The signals from the bloodstream chemistry and the ancient C and A delta fiber system are then assembled in the brain into a series of body maps, or representations of the state of the body. The signals from the external senses are likewise processed into representations of the external environment. These maps of the body state are used to regulate the homeostasis of the body and to elicit behavior. This complex body state mapping occurs entirely under the threshold of consciousness. This system is present in virtually the same form in most mammals and is sufficient to carry out the evolutionary instructions inherent to the organism. In humans, however, this system further represents Damasio's proto-self, because the maps it generates become the raw material for the core self.

The Core Self

To understand the core self, and the core consciousness of which it is a part, it is necessary to imagine the components of a simple self system. To have a self, there must also be an other, or "not self." There must be a knower and an object to be known. According to Damasio, the knower in this case is the organism, and the object to be known is anything arising in the organism's perception, whether external, internal, or in memory. For these two components (the knower and the object to be known) to comprise a complete system, however, they must also have some kind of relationship to one another, they must interact in some way. The knower must be affected by the object to be known. In an organism, this interaction takes the form of any modification of the proto-self by the object to be known. For example, seeing a mouse (the object to be known) may cause a cat (the knower) to feel hunger (a modification of the internal state).

This simple model of a self system illuminates the structure of the core self in a human being. Whenever an object to be known triggers a modification of the proto-self (i.e. affects the internal condition of the body), core consciousness generates a knower relating to an object to be known. *This knower, the core self, is the transient protagonist of the nonverbal "story" that core consciousness generates.* The core self is composed of a second order level of maps, created from the first order maps of body state created by the proto-self. Because these body maps are "ready made," core consciousness can copy them to build its own set of maps of the knower, or core self. This means that the core self is utterly dependent upon the proto-self, and that the core self could in fact be defined as a mental representation of the proto-self being modified by an object.

For each object that modifies the proto-self, a representation of the changed proto-self is

generated. This representation is the core self, the protagonist of a kind of nonverbal story that consciousness is telling itself about its interactions with the world. Each core self lasts only a few seconds, as each "pulse," or sequence of interaction, unfolds. Because of the constant availability of objects to be known, however, a new core self is constantly being generated and so appears continuous over time.

It is interesting to speculate about the evolutionary adaptive value of core consciousness. Why bother to generate the self at all, especially considering the amount of extra brain processing power required? What good does it do to know that we are knowing? Damasio argues that the first result of core consciousness in a wakeful organism is *more wakefulness*. Because the first order maps of the proto-self have been, in effect, duplicated in the core self, the intensity of awareness of those maps is also increased. The second result of core consciousness is that more *focused attention* is given to the causative object. This again is the result of the duplication of the first order maps. As Damasio puts it, the message implied in core consciousness is "focused attention must be paid to X."[22] This increased wakefulness and focus allows for a greatly *enhanced processing* of the object and responses to it. In this way the core self makes evolutionary sense and pays its own way for the brain power such processing requires. In short, having a self allows organisms to respond more adaptively to their environment by being more alert, focused, and detailed.

Core consciousness, then, brings together the second order representations of the knower (the core self), the object to be known, and the relationship between the two. Humans (and a number of higher mammals) have core consciousness and experience a core self. It is the "me" of the moment, the embodied self to which everything is happening. In humans it is the fundamental layer of the conscious self, and yet it is completely embodied, nonverbal, transient, and generic. This is hardly what we usually mean when talking about self awareness or human consciousness. It is not until we get to extended consciousness and the autobiographical self that we become aware of a unique, individual person who exists over time.

The Autobiographical Self

Just as the core self arises in core consciousness, the autobiographical self arises in extended consciousness. One way to understand extended consciousness is that it adds the factor of *time* to core consciousness. Here the core self's "me of the moment" is surrounded and supported by a sea of facts about that same self in the past, and its likely condition and activities in the future.

For example, if core consciousness allows you to know that you are seeing a dog now, extended consciousness allows you to know that the dog you are seeing is your dog, a Rottweiler, named Caesar, that you have raised from a puppy. You can know that Caesar needs to go for a walk now, like he does every day, and that during this walk you will likely stop off at your sister Sue's house where her child Sarah will enjoy playing with Caesar for a few moments, and so on.

Extended consciousness is composed of memories, an enormous database of facts about

everything the individual has encountered (a notion explored at length by LeDoux, 2003) It is also composed of the inferences about the future it can make based on these memories. Extended consciousness permits the view of vast time scales and incredible complexities, a truly rich panorama many orders of magnitude more complex than core consciousness.

The autobiographical self arises in extended consciousness when these autobiographical memories are "fed back into" core consciousness, so to speak. These autobiographical memories can be called up and processed as objects in core consciousness, giving a person a sense of self-knowing. This warehouse of autobiographical knowledge being constantly recalled and illuminated by core consciousness comprises the autobiographical self. This process is happening contemporaneously with the processing of other non-self objects, so there is the continuous sense of "me knowing an object." Extended consciousness furthermore has the processing power to simultaneously hold images of the autobiographical self and the other current non-self objects together in consciousness, and these are "bathed in the feeling of knowing that arises in core consciousness."

Damasio's model also includes a substantial and neurologically detailed account of the arising of the human sense of self from the body upward (although that is not presented here). One of the surprising discoveries here is that the parts of the brain involved in mapping the body state also turn out to be crucial to feelings of emotion. FMRI studies show these brain structures (the insular cortices) are active in the feeling of such emotions as fear, anger, happiness, sadness, and so on. The insular cortices also "light up" when there are feelings of pain, coolness, heat, itch, disgust, sexual arousal, and certain drug highs, "emphasizing the point that this region thoroughly relates to bodily state." Thus *both body state and emotions* come together here to form the proto-self, the basis of the sense of self in a human.

Many animals have the first leg of this system (going as far as the hypothalamus), but only in primates does the system continue on to the insular cortices. As Damasio puts it: "This suggests that while many species can have a continuous representation of the body capable of supporting feelings of emotion and a sense of self, only primates might, through the addition of structures that facilitate high-level convergence, generate the sort of higher-order mappings that would make the sense of self most encompassing."[23] He believes that the C and A delta fiber system, the vagus nerve, the insular cortices, as well as the somatosensory cortex I and II combine to form a continuous map of the body state, which includes input about the internal milieu, the viscera, and the invariant aspects of the musculoskeletalsystem, and this map is the "neural foundation for the self, the grounding of the material 'me'."[24]

Another researcher who sees the body as the foundation for the self is Rodolfo LLinás, although he approaches it from the perspective of movement. For an animal to be able to move, he argues, it must also have an ability to predict the outcome of its movements. That is, it must estimate the effect of its motor efforts before making them. According to LLinás, this predictive ability is the dominant function of any brain, and the centralization of this prediction is the self. According to his research, the physiological location of this predictive ability is the thalamocortical system, which connects the

vestibular nucleus and the cortex. LLinás calls this system the “predictive organ” and emphasizes that it must be unitary, because any organism with more than one predictive center would be chaotic. LLinás also agrees with the CMS notion that the self is merely a representation of a distributed, parallel apparatus, having no true center.

Thus several researchers are looking at how the self arises from the need for the brain to be aware of the body, its state, its emotions, and its mobility. Organisms require at least a minimal implicit self to survive and thrive in the world, and this proto-self is focused on the body and its emotions. While it is not conscious, this embodied, emotional self serves as the foundation upon which the conscious self of higher animals is constructed. In subsequent sections we will repeatedly see how the emotions serve to coordinate and unify the self.

The Mnemonic / Fictional Self

The self has often been associated with memory. As Scottish philosopher James Mill wrote in 1829:

> "The phenomenon of Self and that of Memory are merely two sides of the same fact, or two different modes of viewing the same fact … This succession of feelings, which I call my memory of the past, is that by which I distinguish my Self."[25]

However the actual mechanisms of memory in the brain have been notoriously difficult to determine. As recently as 1950, psychologist Karl Lashley facetiously declared that learning was impossible, given what he had discovered – or failed to discover – in his life long research into memory.[26] Work with brain-damaged patients through the decade of the 1950s, however, began to reveal some of the basics of memory, such as the important role played by the hippocampus.

By the 1990s, researchers had made enormous strides in understanding how memory works. Nobel prizewinner Eric Kandel, in particular, determined the neurophysiological and neurochemical mechanisms that underpin short and long term memory.

As we get a clearer view of how memory works, what does this tell us about the self? And what role does memory play if the self is, as the CMS suggests, fluid and somewhat fictionalized?

Neuroscientist Joseph LeDoux states it flat out: "In the end, then, the self is essentially a memory, or more accurately, a set of memories." [27] To arrive at this statement, he takes as his starting point the synapses of the brain. If the self arises from the brain, and the brain is composed of synapses, then QED the self is nothing other than the synapses.

Synapses are the connections between neurons. The brain encodes information as a pattern of synapses between the neurons engaged in processing an experience. The sum of all these synaptic patterns of coded information about the organism's experience is, in LeDoux's view, "the [key] to who that person is." He calls the model based on this idea the "synaptic self." Another word for these encoded information patterns in the brain is memory.

It turns out that this simple statement requires some qualifying before it completely makes sense. The biggest qualification relates to the brain systems that are not memory-based. Many structures that seem to play a role in the sense of self are determined by genetics, and do not change their function based on changes in memory. LeDoux finesses this question by explaining that genes are, in one way of looking at it, a kind of memory. Genes are the encoded evolutionary memory of our species, and it is the genes that determine the synaptic connections in more fixed brain structures. As LeDoux puts it "…both genes and experiences have their effects on our minds and behavioral reactions by shaping the way synapses are formed."[28]

This way of seeing the self illuminates its distributed qualities. The brain does not store or recall memory in a single, unified way. Many different brain systems process, store, and recall memory locally, meaning that various remembered aspects of a single experience may be found in synapses in widely separate brain regions. Thus if the self is memory, as LeDoux asserts, the self is found peppered throughout the brain.

LeDoux's synaptic self also contributes to the idea that the majority of the self is unconscious. Many if not most of the brain systems involved in memory function implicitly, beneath conscious awareness. Only a small fraction of the memory about an experience is stored in a system that allows for conscious recall. If the self is composed of synaptic connections (memory), and most of those memories are unconscious, it stands to reason that the majority of the self is unconscious.

Furthermore, if most of the self is implicit, based in the synaptic structures of the brain, this makes sense in terms of evolution. The decentralized, largely unconscious self is something that could have arisen in various animals bit by bit over evolutionary time. Earlier views on the nature of the self were based on a large evolutionary gap between humans and animals in terms of the self. Peter Strawson (1959), for example, distinguished material things that are conscious and those that are not. Yet his examples of statements about material things that are conscious ("is in pain," "is thinking," "believes in God") leaves no room between humans and, say, rocks. Animals in this view are no different from dirt in terms of their consciousness. LeDoux makes the point (common in the CMS) that the kind of self humans exhibit must have evolved from something similar in animals. As he puts it "The existence of a self thus comes with the territory of being an animal. All animals, in other words, have a self, regardless of whether they have the capacity for self-awareness. These differences within organisms (conscious vs. unconscious aspects) and between organisms (creatures with and without consciousness) are not captured by an undifferentiated notion of the self, but can be accounted for by recognizing the self as a multifaceted entity, consisting of both explicit (conscious) and implicit (unconscious) aspects."[29]

Another notion that LeDoux explores is how this distributed self can maintain coherence over time, which he calls the "paradox of parallel plasticity." If pieces of the self exist in so many different locations in the brain and each of these systems is processing and storing information independently, there would seem to be a danger of these different systems getting out of sync with each other and the self losing its sense of unity. What keeps the self from fracturing into many separate entities, each with its own competing agenda?

LeDoux postulates four reasons that this fragmentation does not occur in healthy individuals. First, although the systems are involved in different functions, they are all processing the same experiences in the external world.

Second, none of these systems is physically separate from the others. The synapses of any system are connected to the synapses of others throughout the brain. This matrix of connected synapses allows for these systems to interact and coordinate learning and behavior.

Third, there are regions of the brain, "convergence zones," which function specifically to integrate the information of the different systems. Convergence zones not only collate the information coming from different systems, but can also coordinate and even control these systems. They provide a good deal of coherence, and the work of Gazzaniga (and others) is largely about this aspect of the self, as we will see.

Fourth, and most important, there are the chemical systems that affect the brain, particularly those related to *emotion*. Emotion is one of the main ways the self is "glued" together. Emotionally charged events release chemicals throughout the brain that affect learning in a global manner.

The brain has many different emotion systems, such as networks to identify threats and respond to them, or to find food or sex partners. Although these systems have goals that conflict with the others, they do not increase the fragmentation of the self, but rather help to hold it together because when one of these emotion systems is active, the others are inhibited.

For example, under normal circumstances, when an animal is not hungry or excited sexually, its fear of predators will keep it from venturing out of its safe zone. These other needs are inhibited by the dominant emotion of fear. However when hunger arises, the desire to eat will prevail, and the animal's fear of predators will be inhibited. It will then leave its safe zone and venture forth to find food, even in dangerous areas. The same inhibition of fear occurs when the sexual system is active. "People risk all sorts of adverse consequences for a sexual fling," as LeDoux notes. In each case, the activation of a strong emotion in one of these systems pervades the brain, inhibiting the functioning of competing systems. This is one way emotions play a large role in the sense of self.

Conversely, when emotional systems become erratic, as in schizophrenia, the sense of a coherent self can begin to disintegrate. This can become so strong that competing brain systems are heard as the voices of others, seemingly completely outside the self. When the brain becomes overwhelmingly focused on one emotion, as in anxiety disorders, depression, or addiction, the sense of self can become so distorted that an individual seems "like a different person" both to themselves and others.

The emotions, then, play a strong role in the sense of self, even if this self is located in the memory (and structure) of the brain. Yet memory is, after all, nothing more than the storage and retrieval of information in the brain. How accurate is this storage/retrieval mechanism, and what implications does this have for the self? Schachter, et al, in his paper The *Seven Sins of Memory* looks at several "flaws" or types of inaccuracies in memory that actually help to strengthen the ego.

One type of inaccuracy in memory is called *misattribution*. This means that memories are assigned to an incorrect source, for example, when a fantasy is taken for an actual event in the past. Misattribution can seriously affect the sense of who we are as people. Shachter gives an extreme example of misattribution in the case of a patient, H.W. who had an aneurysm and lost memories of previous events. To patch up the holes in his

memory, H.W. simply fictionalized his past whole hog. Although he had been married for over 30 years and had four children, he claimed to only have been married for four months. Despite the absurdity of the claim ("Not bad for four months!" he said about his children), H.W. was completely convinced of the reality of his statements:

Moscovitch: I think when I looked at your record it said that you've been married for over 30 years. Does that sound more reasonable to you if I told you that?

H.W.: No.

Moscovitch: Do you really believe you have been married for four months?

H.W.: Yes.

H.W. essentially constructed a false self from misattributed memories. His implicit self took a fantasy and assigned it to memory in an effort to shore up his missing past. The example of a brain-damaged patient is extreme, but a similar mechanism is at work in the average brain. Schachter gives the example of the DRM paradigm (Roediger & McDermott, 1995), in which subjects are given a long list of words related to a missing theme word. When then asked to recall the list, a high percentage of participants will include the theme word, and indicate a great deal of confidence that they saw this word on the list. They are convinced they remember something that never actually happened. The DRM paradigm relies on the close association of the them word to the list, and it is clear that the participants have correctly remembered the *gist* of the word list. Schachter has done several studies that suggest the hippocampus is involved in this "gist memory" and that this may play an important role in the creation of the self from memory.

Another type of error with implications for the self is called *egocentric bias*, the tendency in humans to recall past experiences in a way that make them look better. Several studies (such as Ross & Wilson, 2000) have shown the numerous ways that past memories are re-shaped or fictionalized to enhance the sense of self. Other researchers, such as Rogers (with Kuiper & Kirker, 1977, and since extended by many others), have demonstrated that self-referential memories are better remembered than other types of memories. This suggests that the self wields a strong influence over what we encode as memory and how we will later recall that memory. Schachter suggests that self-referential memory involves a qualitatively different type of encoding than other semantic information. *Self-referential memory is probably not a stronger type of memory, but actually a different type of memory involving different brain structures*. [30] Thus the common practice of telling "tall tales" that inflate one's role and importance in past events has a basis in neurology, and probably reflects the work of the brain to keep the self strong and coherent.

A third memory error relevant here is called *consistency bias*, which means to recall past opinions and actions in a way that is more consistent with their present opinions and actions than they actually were. A striking example of this involves a study (Marcus, 1986) on political opinion. In this study, people rated their opinions on political issues, once in 1973 and then again in 1982. In 1982, they were also asked to remember what their opinions were in 1973. People misremembered their previous opinions in a way that much more closely matched their current beliefs. This seems to indicate that the

memories of self are selectively edited to give an exaggerated picture of consistence and coherence over time, thereby reinforcing the self.

Thus the self can not only be thought of as existing in memory, but the memory systems themselves continuously work to enhance the self by distorting or fictionalizing memories.

Among all the unconscious fictionalizations the self generates, perhaps the greatest is the sense of authorship, that is, the notion of *conscious will*. Harvard psychologist Daniel Wegner has developed the "theory of apparent mental causation" to describe this particular self-deception.

Our minds, says Wegner,[31] are interesting because they not only look out onto the world and see what's going on, they also provide us with "views of *themselves*." The mind creates its own sense of how it works, it's own "self-portrait" which seems, from the view of the person experiencing it, to be complete and convincing. Yet this mental self-portrait is at best a drastically simplified model of what is actually going on in the brain. For example, if you pick up a cup, the mind's explanation is "I wanted to pick up the cup, so I did." There is a thought about doing something and then the doing of that something occurs. This explanation dismisses the tremendous amount of work done by unconscious brain systems to make that simple thought and action possible. In Wegner's metaphor, the mind's explanation of itself ignores all the "machinery in the mind's basement that might be creating this conscious show." You think of picking up a coffee cup and then pick up that cup "not because thinking causes doing, but because other forces of mind and brain (that are not consciously perceived) cause both the thinking and the doing."

The self-portrait of the mind, then, is virtually a cartoon compared to what is really going on. Yet this idea that we will actions consciously (the sense of authorship) is so convincing that it has had lasting ramifications in psychology and neuroscience long after the notion has become scientifically untenable. Wegner has therefore attempted to investigate exactly how this sense of authorship arises. How do we get the idea that we consciously cause an action?

In short, we *infer* it. It is an intuition that occurs because all of the unconscious processes underneath thinking and acting are invisible to the mind's self portrait. It cannot see, or ignores, all that data and only looks at the thinking and the doing.

In a series of experiments, Wegner has shown that in order for the mind to attribute authorship to an action (i.e. the sense that it willed an action to occur) three conditions are necessary: *consistency*, *priority*, and *exclusivity* of thought. That is, "the thought must be consistent with the action, occur just before the action, and not be accompanied by other potential causes." There may be many other thoughts about other things occurring at the same time, but these do not give us the sense of willing the action. They seem irrelevant. There also may be thoughts that *are* relevant to the action, but if they occur too far in advance of, or too long after, taking the action, they also do not give rise to the sense of

authorship. And if an action that we take appears to be caused by another agent, that too disallows apparent mental causation.[32]

Consistency

Patients who received electrical stimulation of the brain in the motor cortex (Penfield, 1975) were "shocked" into producing complex, multistage, voluntary-seeming movements. Externally these movements appeared to be voluntary, yet the patients said that they had not done the action, and that Penfield had "pulled it out" of them. Because they had no thoughts consistent with the action that they were induced to make, it felt unwilled to them.

In one of Wegner's own studies (with Gibson, 2003), participants were indirectly exposed to a prime word, *deer.* They were then asked to type letters randomly on a keyboard without seeing what they typed. They were then given a list of the words that a computer had supposedly extracted from their random typing, when in fact none of the words had been typed. Asked to rate their feelings of authorship for the words on this list, participants consistently reported higher authorship ratings for the word *deer*, and even the related word *doe*. This finding suggests that when people have prior thoughts consistent with an action – *even one they never actually performed* – they can feel that they willed that action.

Priority

Wegner (with Wheatley, 1999) also devised an experiment to study the timing of consistent thoughts relevant to an action. Participants were presented with the sound of the word *swan* (the thought) while moving an computer cursor to a picture of a *swan* (the action). Unknown to them, the movements of the cursor were actually being controlled by an experimenter who "gently forced the action." Regardless of the fact that they were not actually making the action, participants exposed to the relevant thought (the word *swan*) from 1 to 5 seconds before the action felt that they had acted intentionally. If the thought was provided 30 seconds in advance of the action, or 1 second afterward, they did not assign authorship to the action. So even when a person does not actually take an action, and the thought of that action occurs in the crude form of a word in a set of headphones, *the mere timing of the thought* causes people to believe they willfully took the action.

Exclusivity

The presence of competing possible authors for an action can strongly undermine the sense of conscious will for that action. In another experiment, Wegner (with Dijksterhuis, Aarts, & Preston, 2003) created a situation where participants were asked to guess whether they or a computer had caused an action, under conditions in which it was impossible to know for sure. If the participants were subconsciously primed with the words "I" or "me" (flashed extremely quickly on a screen) they were more likely to attribute the authorship of the ambiguous action to themselves. If, on the other hand, the word "computer" was very briefly flashed onscreen, participants were less likely to claim authorship. And when they were primed with the name of a third-party actor, "God," they

were also less likely to claim authorship. Thus even when the competing possible agents for an action are known only subconsciously, their presence seriously undermines the sense of having willed that action. This finding is supported by other robust studies.[33]

Wegner speculates on the possible reasons for the brain's habit of assigning apparent mental causation, even when it is inaccurate. He hypothesizes that the mind's self portrait may necessarily be very limited due to *capacity constraints*; it would just take too much processing power to create a more complex and accurate picture of the situation. Secondly, self-insight may be shaped by "*theory of mind*." The same simplified mechanism that evolved to give humans insight into the minds of others is also directed at themselves. Finally, Wegner suggests that the presence of *conscious previews* of an action strongly imply causation, and the fact that our mind's create such previews gives rise to a enhanced sense of causation.

Wegner also discusses the evolutionary advantage of the sense of conscious will by postulating an *authorship emotion*, the feeling of responsibility for an action. Feelings of authorship give rise to moral emotions such as pride and guilt, and therefore serve as "somatic markers" that bring attention to evolutionarily relevant behaviors (as Damasio suggests, 1994). As a person subscribes actions to themselves over time, they begin to see themselves as a certain kind of author, a person who does good or bad things. For our purposes it is interesting to note that, in the end, the authorship emotion serves to strengthen the sense of self. As Wegner puts it, "the self can be understood as a system that arises from the experience of authorship, and is developed over time. We become selves by experiencing what we do, and this experience then informs the processes that determine what we will do next. The self, in this view, is not an agent, an origin of action—but instead is an accumulated structure of knowledge about *what this particular mind can do*."

Besides creating the sense of authorship, the brain has other mechanisms to unify the sense of self, and one of these, proposed by psychologist Michael Gazzaniga, is the "left brain interpreter module." Gazzaniga's work with split brain patients has shown the self to be highly distributed throughout the brain. This complex, multifaceted network of different neuronal systems requires some mechanism for maintaining the sense of unity over time. There must be some system that takes the work of many competing modules in the brain that process semantic input and produces a single, coherent narrative.

Studies conducted by his colleagues (Klein, 2002; Kelley, 2002) and others seem to indicate that processing of information about the self is unique from other types of semantic processing. This special self-processing, however, is widely distributed throughout the brain. Research suggests that self-processing in each hemisphere can function largely autonomously—as if each brain has its own self. Yet this situation does not devolve into a fragmentation or fracturing of the unified sense of self, says Gazzaniga,[34] because of the presence in the left brain of an "interpreter" module that unifies the input of both hemispheres into a single, narrative whole.

Fairly robust data supports this conclusion. Gazzaniga's work with split brain patients gives fascinating insights into the nature of this left-brain interpreter. The left brain's interpretive module commonly invents stories to explain what the right brain is experiencing. In a classic study (with LeDoux, 1978) he presented corpus callosotomy patients (i.e. people who have had their cerebral hemispheres surgically cut off from each other, resulting in a "split brain") with different images to different hemispheres. For example, an image of a chicken claw was shown to the left hemisphere and an image of a snow scene to the right hemisphere of subject P.S. The subject then picked a picture of a shovel with the left hand and a picture of a chicken with the right hand (of course the sides were reversed). This is logical, because a shovel can clear snow and a chicken claw goes with a chicken. Thus both hemispheres were choosing correctly, based on the information they were receiving. Explaining these choices, however, P.S. said that "The chicken claw goes with the chicken, and you need a shovel to clean a chicken shed." This indicates that the left hemisphere interprets the experience of the right hemisphere in terms of the left hemisphere's own viewpoint. With a wealth of such cases, Gazzaniga makes a strong case that the left-hemisphere interpreter functions to unify the narrative of the two hemispheres, even when it has no access to information from the right hemisphere. The emphasis of this module is on creating a coherent narrative, no matter how illogical or fictional this narrative is.

Thus Gazzaniga's work not only supports the idea that the self is widely distributed, but explains the unification function in terms of a left-hemisphere structure that generates narrative cohesion, no matter the cost to veracity.

The Developmental / Evolutionary Self

The distributed, multilayered self is a complex construction which does not arrive ready-made in a newborn, but develops over many years as children grow into adults. This process gives us the opportunity to learn about the self by carefully observing it "under construction." We can also gain insight into the self under construction by uncovering its rise in hominids over the course of evolution.

Child psychologist Dr. Michael Lewis traces the growth of the conscious self in children, and finds the explicit self appearing at around 15-18 months of age. He postulates that children are born with what he calls the "machinery of the self," a basic, unconscious understanding of the difference between self and other. This corresponds roughly to Damasio's proto-self. The machinery of self is completely biological in nature (i.e. not learned), and exists at a rudimentary level even in organic systems as simple as T-cells. This basic underlying system for selfhood in humans is "…an elaborate complex of machinery that controls much of our behavior, learns from experience, has states and affects, and affects our bodies, most likely including what and how we think."[35] Importantly for our discussion, the machinery of the self provides the basis for the explicit self, what Lewis calls the "idea of me."

Lewis has assembled his work and that of others into a theory of development of the conscious self. In his theory, the machinery of the self eventually grows to become the implicit self and the idea of me grows into the explicit self. Lewis postulates that there are three salient features that mark this arising of the explicit self, all of which occur around the second half of the second year: self recognition, the use of the personal pronoun, and pretend play.

Self Recognition, Personal Pronouns, and Pretend Play

Clearly one of the first things we would imagine belonging to a sense of self would be the ability to recognize oneself consciously. Studies of *self recognition* (Lewis, 2003; Lewis & Brooks-Gunn, 1979b, Lewis & Ramsay, 1997) typically use the well known "mark test," in which a mark of some kind (such as a spot of lipstick) is put on the nose, where it can only be seen in a mirror. Infants begin very early (as young as 2 months) to interact with their image in the mirror, although there is every indication that they treat the image as another child. They smile at it, coo, touch it in the mirror, and – when their motor skills come online – attempt to crawl behind the mirror to find this mysterious "other" they see there. Between 15-18 months, however, children's behavior toward their mirror image changes radically. They see the mark on their nose in the mirror and now touch their own nose, or make a comment about their nose. This is taken (as it is in similar animal studies) to indicate that they now have the concept of "me." The explicit self has arrived. It is interesting to note that several animals besides humans can "pass" the mark test, including chimpanzees and orangutans, and possibly elephants, dolphins, and other creatures (although these latter outcomes are questionable), suggesting that they, too, have at least the rudiments of a conscious self.

The second feature of the emerging explicit self is the use of *personal pronouns* such as "I", "me", and "mine." Much research into the development of the self in children is hampered by the fact that the studies are dependant on complex language skills. They may only be measuring the child's capacity to describe what the child is experiencing, rather than actually demonstrating the arising of the conscious self. The use of personal pronouns is relatively immune to this problem (due to their non-complexity), and is considered by scientists to be a mark of emerging self awareness. This conclusion is reinforced by the behavior of children when they use personal pronouns. Typically they will say "me" or "mine" when pulling an object away from another child towards themselves, for example, a clear indication that they understand the pronoun refers to themselves. According to Harter (1983) and Hobson (1990) the use of personal pronouns begins in the second half of the second year and demonstrates the child now has self-awareness. Thus the appearance of this second feature fits occurs within the same time span as the first.

The third feature signaling the development of the conscious self is *pretend play*. Many theorists[36] have shown that pretense is the result of the capacity to understand one's mental state. Pretense requires a knowledge of "who I am" to be able to also hold the idea of "who I am pretending to be." Significantly, this co-arises with the rudiments of a theory of mind in children. Pretend play begins around the same time as the other two features (15-20 months).

Just as we can observe the self under construction in the course of a single human life, we can also see it growing more complex in the course of evolution. Professor of psychology and neuroscience at Duke University, Mark Leary, (with Buttermore, 2003) has looked back at the fossil record of proto-humanity in an attempt to theoretically reconstruct when the modern self came into being. He ascribes to the overarching notion that the self is not unitary, but instead composed of many distributed cognitive modules, each of which evolved at different times to fulfill certain goals for the organism. His analysis of the evolution of human self-consciousness relies on Neisser's (1988, 1997) model of self-knowledge that consists of five unique types of information: *ecological, interpersonal, extended, private*, and *conception*. These types of self-knowledge arise at different times in child development, and are present or absent to varying degrees across species, suggesting that they may have arisen at different times in human evolution. Human beings are the only animals to possess all five types of self knowledge. These are:

> *Ecological-self ability* – allows an organism to process information about its physical environment.
>
> *Interpersonal-self ability* – allows an organism to process information about its "unreflected" social interactions with other members of its species. ("Unreflected" means that it is not conscious)
>
> *Extended-self ability* – allows an organism to reflect on itself over time.

> *Private-self ability* – allows an organism to process private, subjective information such as thoughts, feelings, intentions, images, etc.
>
> *Conceptual-self ability* – allows an organism to process abstract, symbolic representations about itself.

Leary suggests that these abilities do not evolve only qualitatively but also quantitatively, so that they may be present to different degrees in different organisms. For example, both humans and chimps have the capacity to think about themselves in the future (Extended-self ability), but this ability is developed to a much higher degree in humans, allowing them to look much further into the future. Thus when looking at the fossil record, we can see certain of these abilities getting stronger over time.

[Leary goes through all the hominids, painstakingly reconstructing each little speculative bit of self-consciousness. See diagram in Appendix]

For most of *Homo*'s history, things changed extremely slowly and very little. As Leary's analysis shows, sometimes there is no discernable change in self-knowledge over periods sometimes as long as a million years. But during the late Paleolithic era (about 40,000-60,000 years ago) specialized stone tools and recognizable features of human culture appear suddenly. This period shows such a precipitous and dramatic shift that it is known to archeologists by a series of incendiary names, such as the "great leap forward" (Diamond, 1992), "cultural big bang" (Mithen, 1996), "creative explosion" (Pfeiffer, 1982), "cultural explosion" (Boyer, 2000), "human revolution" (Deacon, 1989), and "dawn of human culture" (Klein & Edgar, 2002).

Proto-humans before this period were so different from modern humans that we probably cannot imagine what their lives were like. It is not until the cultural big bang (the term Leary uses) that they become recognizably human to us. There have been numerous attempts to explain why the cultural big bang occurred – and they all have some merit – but Leary feels that any discussion of this is incomplete without including the idea that the *human self changed drastically at this time*.

Leary argues that the missing piece of the human puzzle up until this point was *conceptual-self ability* (number five in Neisser's model), and that the cultural big bang was caused by the arising of the conceptual self. "Combined with the extended-self ability that allowed them to project themselves in the future, this new ability allowed people *to imagine themselves in the future in symbolic and abstract ways*, a trait needed for intentional innovation, symbolic culture, and efforts at self-improvement."

There are several areas in which Leary sees the mark of this new conceptual-self ability. For example, human technology undergoes a veritable tsunami of change and innovation at this time. Tools, shelter, and clothing all show remarkable advancement and variation at this time, and boats are invented. These changes required a "forward-looking self that could imagine novel solutions to life's challenges," i.e. the conceptual-self. The arrival at this time of art, body adornment, and ritual burial also point to a self that can think about itself in symbolic ways.[37]

Leary's concept adds nuances to ideas about the evolution of the human self. It explains it as a slow, gradual layering of new abilities to the self found in animals, and yet sheds light on the reasons for the sudden explosion of creativity in the cultural big bang. Furthermore, it is interesting to note that although the cultural big bang occurred about 50,000 years ago, anatomically modern humans arrived perhaps 100,000 years *before that*. This suggests that *the differences in neurology that brought about the conceptual-self ability may be very minute*, too small to show up in fossils. While there are problems with this notion (it would require two separate dispersions of humans out of Africa or a simultaneous arising of the ability in disparate populations), it fits well with the CMS concept that the self is distributed among many different cognitive modules in the brain, each of which arose separately in evolution, and most of which are implicit.

The Socio-Cultural Self

Human beings do not live in isolation, rather they are commonly found in large social groups having distinct cultural qualities. This socio-cultural environment affects the construction of the self in childhood, and continues to influence the adult self. This influence appears to be mainly implicit and can demonstrably change an individual's self-concept from moment to moment.

Mahzarin Banaji has shown that implicit cultural and social attitudes towards group identity strongly shape our beliefs and behavior *even when they are in direct contradiction to explicit beliefs.* For example, Banaji (with Nosek) subjected liberal college students to Implicit Association Tests, which forced them to respond to stimuli presented too briefly for conscious appraisal. Although students claimed to have no prejudices, the results of many thousands of tests show that they have the same cultural biases as everyone else, albeit on the implicit level. Thus group identity – whom I am socially – exists partially on the unconscious level, where it is immune to the higher values I may consciously profess.

There is a robust series of Implicit Association test results that show a strong association of self-identity and group association. Banaji (and Devos, 2003) looked at the strength of implicit national identity in US citizens.[38] Participants were shown symbols or words associated with America or foreign countries (flags, coins, monuments, maps) and combined with pronouns indicating ingroup ("we," "ourselves," etc.) and outgroup ("they," "other," etc.). Participants could complete the task much more quickly when American symbols were associated with the ingroup words rather than the outgroup words, suggesting that the identification with national group was automatic and therefore quicker. Banaji working with others has generated similar results for sexual identity, and even college association. Although earlier theories assumed that the learning of such implicit identities happened slowly, Banaji's studies show such ingroup identities forming very fast. For example, it only took participants 3 to 4 days to reach the same level of identification with their college as those who had been on campus 3 to 4 years.

Thus the implicit representation of the self is partially composed of ingroup identity. Banaji asserts that "group membership comes to be automatically associated with the self and that people automatically endorse attributes stereotypic of their group as also being self-descriptive." Again this appears to be unconscious and automatic.

Citing a large number of studies by various researchers, Banaji demonstrates that people show a large bias towards ingroup members across a wide spectrum of possible ingroups. For example, white, rich, American, straight, Christian students showed a strong preference for people who were white (rather than black), rich (not poor), American (not foreign), straight (rather than gay), and Christian (rather than Jewish) (Cunningham, Nezlek, and Banaji, 2001). There is similar data for Japanese and Korean Americans. Each of these ingroup preferences develops in combination with the others, which Banaji feels is evidence for an "implicit ethnocentrism dimension." Other studies use linguistic structures to show these same biases.

These biases develop under the most minimal conditions. Perdue, Dovidio, Gurtman, and Tyler (1990) found that when participants were implicitly primed with ingroup pronouns ("we," "us," etc.) they responded to pleasant words faster than when primed with outgroup pronouns ("they," "them," etc.). Otten and Wentura (1999) showed that otherwise neutral words would immediately acquire a positive affect when associated with an ingroup, and when associated with an outgroup, neutral words acquired a negative affect. Thus the mere presence of ingroup / outgroup association is sufficient to create an implicit bias about an otherwise neutral term.

But it goes further than this: that the "mere ownership of an object or its association to the self is a condition sufficient enough to enhance its attractiveness." For example, Nuttin (1985) showed that individuals asked to choose letters from a random list consistently choose letters from their own name. This is known as the "name letter effect" and has been shown to exist in many countries. Nuttin further showed (1985) that participants were unable to find a meaningful pattern in the letters they had selected, suggesting that they were *consciously unaware of the reason behind their choice*. This research, combined with other studies showing, for example, a similar preference for numbers associated with ones birthday, or people having the same birthday, having names beginning with the same letter, living in the same city, or having the same careers, seems to indicate an unconscious bias towards the self in attitudes. Further experiments (Feys 1991) with icons randomly assigned to either the participant or a computer, found that participants consistently evaluated their "own" icons as aesthetically more attractive. Again, the simple act of owning an object, no matter how tenuously, endows that object with increased value due to implicit self-esteem. The identification of the self with an object makes that object more valuable, and this identification happens on an entirely unconscious level.

These unconscious biases can strongly influence our self-evaluations, behavior, and motivations. In terms of self-evaluation, subtle priming can affect an individual's self-evalutation. S. Sinclair, Hardin, and Lowery (2001) unobtrusively asked participants to report either their gender or their ethnicity. They then had them indicate how others and they themselves evaluated their verbal and math abilities. Despite (or perhaps because of) the unconscious nature of the manipulation, it influenced evaluation. For example, when gender identity was triggered, Asian American women said that others were more likely to evaluate them as having higher verbal skills than math skills. When ethnicity was triggered, they felt that others would see them as having higher math skills. Significantly, *they also evaluated themselves in the same way*. The mere presence of an implicit stimulus regarding group identity was enough to change an individual's self-evaluation. This seems to happen much more strongly when it is implicit rather than explicit.

Behavior is also affected by these implicit biases. Shih and colleagues (1999) showed that Asian American women perform better on a math test when ethnic identity is activated, and perform worse when gender identity is activated. Such behavioral changes occur in a wide variety of situations and, in most of these cases, the individuals involved are unaware of any change in behavior.

Banaji also remarks on the existence of both explicit and implicit self-concepts. Several studies have shown that *explicit self-esteem and implicit self-esteem are distinct*

constructs, that can be quite different in nature. Additionally, implicit self-esteem appears to be a much better predictor of behavior than explicit self-esteem (Spaulding & Hardin, 1999). Paulhus and Levitt (1987) furthermore showed that emotional arousal predicts a shift in measures of self-esteem. When emotional distractors are introduced, individuals claimed more positive traits in their self descriptions. Koole and colleagues (2001) also found that emotions triggered stronger implicit self-esteem effects (bias for name letters and birthday numbers again).

Thus it is clear that the sense of self has a strong social component, and that this socially-constructed self is primarily implicit. Banaji remarks on the irony of this, given that "of all the special domains of knowledge, self-knowledge is assumed to be well known and well defended against intrusion. [Studies such as we have seen] both highlight the pervasiveness of social influences on the self and point to the inadequacy of introspection as the only tool for obtaining self-knowledge."

Cultural Anthropologist Naomi Quinn agrees with this characterization, saying that "one of the most prominent features of cultural knowledge of all kinds is that it is overwhelmingly implicitly transmitted." [39]

Certain tasks are routine, important, too complex for individual solution, and widely applicable across all members of the society. Under these conditions a cultural solution becomes attractive, and Quinn argues that child-rearing is the "paramount" example of such a task. Cultures everywhere do not leave such an important task to the individual, but evolve cultural forms of socializing children.

In her look at child rearing practices across several cultures (American, German, Kenyan, Micronesian, and Inuit), Quinn and her colleagues have found three universal features of child socialization: *constancy of experience*, *emotional arousal*, and *moral valuation*. These features are designed to link a child's life lessons with strong emotions (such as fear, shame, etc.) and with their sense being a good or bad person. Despite the many different cultural forms it takes, this lesson-based triggering of emotions and moral self-worth are intended to fundamentally shape the child's implicit self-identity into the kind of person the society deems valuable.

Constancy of experience, says Quinn, means both the consistent reinforcement of the same lessons over and over, as well as the exclusion of contradictory signals. This experiential constancy creates synaptic patterns that are "highly resolved," meaning that they are strongly connected to one another, and not connected to other synapses. Such a highly resolved group of synapses will fire all at once, and not trigger any others, thus generating a powerful, automatic response. Child rearers achieve this kind of regular, exclusive repetition both explicitly, through injunctions, admonitions, and corrections, but also implicitly, through look, gesture, and body language. Because such "embodied," implicit measures are highly habituated, Quinn feels that they "converge to immerse the child in a cultural world of a certain constant shape, conveying their lessons repeatedly, redundantly, and unmistakably." The more these implicit teachings are repeated, the stronger the synaptic connections will be.

For example, Gusii (? – check for accuracy) mothers want calm, quiet children. They will "punish" an unruly child by looking away from it and ignoring it. This lesson, given in the language of the body, may be repeated many times a day for years, teaches the child what kind of behavior is rewarded with attention, and what kind is punished with inattention. The shear number of repetitions of this lesson lays down a solid synaptic "road."

Emotional arousal deeply reinforces these synaptic connections. Parents worldwide beat, frighten, tease, shame, and praise their children, causing them to have strong emotional associations with culturally important lessons. Quinn invokes LeDoux here,[40] mentioning his statements about how emotional arousal causes hormones to be released that organize and co-ordinate brain function (recall LeDoux's notion that emotions unify competing brain modules), and strengthens the highly resolved synapse groups. Lessons imparted with emotional arousal are so deep that children will never forget them, and will often re-enact the lessons thus etching the teaching even deeper synaptically.

Parents wishing to emotionally arouse children commonly frighten them. One technique to accomplish this that is found worldwide is for adults to dress up as spirits, monsters, or other terrifying apparitions. The Ifaluk people, for example, dress up as a special type of ghost which is said to kidnap and devour children. If a child misbehaves, the ghost is summoned from the forest. When it appears, all the children around leap into the arms of their parents, very afraid. One of the parents then tells the ghost that "the child will no longer misbehave" and the ghost can go away. The level of fear that accompanies such a lesson reinforces it dramatically.

Rearers also use *moral valuation* – explicitly labeling the child as good or bad – to arouse the emotions of the child even more completely. The disapproval of a care-giver is emotionally arousing because a child interprets this as a threat to security and survival. Approval, alternately, signals that the child is safe and secure. The arousal of the social emotions of approval and disapproval lock in place the lesson synapses even more fully, creating a person that throughout life will seek approval by acting in a culturally appropriate manner, and avoid the pain of disapproval by foregoing the misbehavior that the culture does not want. A common example of this is the cries of "Good boy!" or "Good girl!" given in a special tone of voice and accompanied by claps that is used in American households to emphasize good behavior.

These three techniques work regardless of what the cultural lessons actually are, and indeed the lessons are quite different from culture to culture. Micronesian children, raised on small, densely populated Ifaluk Island, are strongly socialized to exhibit a trait called *malewelu* (calmness), which means to avoid any behavior that could feel aggressive or disruptive to others. Gusii children from western Kenya are taught to be subordinate, respectful, and obedient, as are Chinese children. German children are imbued with a sense of *Ordnungsliebe* (love of order), self-containment, and self-reliance.

Such values can be split not only along cultural lines, but along the lines of economic class even within the same culture. Working class children in America are taught the value of self-reliance, for example, whereas the lessons of upper class children focus more on self-expression.

The cultural self that early training builds so forcefully in the mind of a child remains in place for a lifetime, implicitly enforcing the values of the culture below the level of conscious awareness. These values remain largely unconscious in the adults of a given culture, and it often takes foreign observers to explicitly state what these values consist of.

Social psychologist Hazel Rose Markus believes this culturally-created self is not superficial, but is in fact a core part of individual identity. The cultural effects of child rearing are so deep and lasting that it is possible to trace these effects in the external actions of adults, where these cultural meanings are expressed and reinforced. The cultural construction of the self builds into each individual a sense of what it means to be a self acting in the world. "Being a person and acting in the world … are culturally saturated processes that entail engagement with culture-specific sets of meanings and practices, what we call *models of agency*"[41] (italics mine).

To illustrate models of agency, Markus focuses on the differences in agency between European Americans and Asians. European Americans tend toward a very individualistic view of agency that Markus calls "disjoint," in which good actions are characterized by being self-focused, independent of others, and freely chosen. Asians, on the other hand, express a model Markus labels "cojoint," in which good actions are characterized by being relationship-focused, interdependent on others, and responsive to obligations and expectations of others.

Markus' lab has conducted numerous experiments which confirm these cultural biases in agency. European Americans will make choices that affirm their individuality, whereas Asians will make choices that affirm how well they fit into the group. In fact the very idea of making a choice is considerably different between the two cultures. To a European American, making a choice is "saturated with meaning," the ultimate statement of individuality, and a serious matter that expresses a person's unique preferences, goals, and convictions. To Asians choices serve no such function. In the cojoint model of agency, choices are opportunities to reinforce appropriate relationship, social status, and obligations. Markus and Kim (1999) presented adults at an international airport with a choice of five pens, four of which were the same color, the fifth being a different color. The participants were asked to choose a pen from among these five that they liked as a gift. 78% of European Americans chose the unique pen, whereas only 31% of Asians did so. Thus the European Americans may have acted to emphasize the culturally important value of expressing their unique individuality, whereas the Asians more often chose to emphasize the culturally significant value of "fitting in."

Markus and Kim (1999) also looked to advertising for evidence, based on the premise that ads show what a culture values and emphasizes. An analysis of American full-page magazine ads found themes of uniqueness, freedom, choice, and rebelling against norms. These ads were organized from the perspective of the actor (i.e. from the view of disjoint agency), urging readers to "be free, declare your independence, think differently, ditch the Joneses, find your own road, be a driver not a passenger, be an original, you can do it" – powerfully emphasizing the idea that the right way to be is independent. A look at

Korean ads, on the other hand, revealed themes of respect for group values, following trends, and group harmony. Korean ads were organized from the perspective of others (i.e. from the view of cojoint agency), urging their readers to "be like us, follow the trend, try to do it in the traditional way, be a good role model" – emphasizing the culturally sanctioned role of the individual as interdependent.

The human self appears to be profoundly influenced by culture, both during childhood, and throughout adulthood. Every action we take reflects who we think we are, and this sense of self arises from an unconscious complex of socially and culturally constructed cognitive schemas. In a very real sense, we are who our culture shapes us to be, and we in turn shape culture by the actions we take in the world.

Conclusion

The self turns out to be a very different thing than we, as humans, have always experienced it to be. It is multiple rather than unitary, more unconscious than conscious, filled with fictions and half-truths rather than the "truth of who we are," contains many elements that are shared with animals rather than the sole possession of human beings, and is rooted in the body and emotions rather than the loftiest, immaterial cognitions. This is the reality to which the Current Model of the Self points.

Science has shown us how memory works on the level of individual neurons, how our brains process many types of information in ways that benefit the self, and how whole cultures affect our sense of self. Now the most far-reaching goal of research on the self is to bring together both ends of the story. That is, to tease out the specific links between how individual synapses function, and how the brain as a whole functions. Some feel that such a goal may be nearly impossibly complex. There are, after all, something like half a quadrillion synapses in the brain of a mature adult. Tracing these connections and understanding how they all work is possibly beyond the ken of even the most diligent science. And yet memory pioneer Eric Kandel suggests, this synthesis is the next step for neuroscience.

Appendix – Leary's (& Butterworth, 2003)

Table 2. Proposed chronology of the development of the human self

Approximate Dates	As Exemplified by	Self Abilities Present[1]	Behavioral Markers
6.5–4.5 mya	Common ancestor of human beings, chimpanzees, and bonobos	Ecological – √√√ Interpersonal – √√√ Extended – √ Private – √ Conceptual – No	Behaviorally resembled modern chimpanzees or bonoboss
4.4–1.2 mya	Australopithecus (several species)	Ecological – √√√ Interpersonal – √√√ Extended – √ Private – √ Conceptual – No	Behaviorally resembled modern bonobos
1.9–1.6+ mya	*Homo habilis*	Ecological – √√√ Interpersonal – √√√ Extended – √√ Private – √ Conceptual – No	Made crude stone tools Carried tools
1.7 mya–1.5mya	*Homo ergaster*	Ecological – √√√ Interpersonal – √√√ Extended – √√ Private – √ Conceptual – No	Improved stone tools
1.2 mya–400,000 ya	*Homo erectus*[2]	Ecological – √√√ Interpersonal – √√√ Extended – √√ Private – √√ Conceptual – No	Improved stone tools Control of fire Cooperative hunting Extensive dispersion
700,000–200,000 ya	*Homo heidelbergensis*	Ecological – √√√ Interpersonal – √√√ Extended – √√ Private – √√ Conceptual – No	Minimal behavioral change over previous species
300,000–35,000 ya	*Homo neanderthalensis*[2]	Ecological – √√√ Interpersonal – √√√ Extended – √√√ Private – √√ Conceptual – No	Crude clothing Care of the elderly Improved stone tools
120,000 ya-present	*Homo sapiens*	Ecological – √√√ Interpersonal – √√√ Extended – √√√ Private – √√√ Conceptual – √√√	Specialized tools Culture, art, music Dwelling construction Boats Ritualistic burial

Notes. [1] √ = appeared to possess a rudimentary self ability
√√ = appeared to possess a self ability with more limited functional capacity than modern human beings
√√√ = appeared to possess a self ability with functional capacity roughly equivalent to modern human beings

[2] *Homo erectus* and *neanderthalensis* were not ancestral to *H. sapiens*. They are included here simply for comparison.

Notes

---

[1] Known in the West by his Latinized name, "Avicenna"

[2] Nahyan A. G. Fancy (2006). *Pulmonary Transit and Bodily Resurrection: The Interaction of Medicine, Philosophy and Religion in the Works of Ibn al-Nafīs* (d. 1288), p. 209-210

[3] Goodman, Lenn Evan (2006). *Avicenna*

[4] LeDoux, Joseph, (2002). *Synaptic Self*, pg. 23

[5] This paragraph on the shortcomings of the cognitive approach draws heavily on Ibid, pg. 24

[6] With the notable exception of the Buddha

[7] LeDoux, Joseph, (2002) *Synaptic Self*

[8] Carl Rogers defined the self as "the organized, consistent, conceptual gestalt composed of the characteristics of "I" or "me."

[9] Ibid.

[10] Hamilton, William D. (1996). *Narrow roads of gene land*

[11] Gazzaniga, Michael (1998). *The Mind's Past*, pg 21

[12] Schacter, Daniel, (2003) *The Seven Sins of Memory*

[13] As we will see in many examples in the course of this paper

[14] Asian American women example from Banaji, Mahzarin & Devos, Thierry (2003). *Implicit Self and Identity*

[15] Damasio, Antonio, (2003). *The Feeling of What Happens*, pg

[16] Damasio, Antonio (2003). *Feelings of Emotion and the Self*

[17] Llinás, Rodolfo (2001). *I of the Vortex*

[18] The structure and to some extent the choice of authors in this paper rely on Henry Moss' *Implicit Selves* (2003), which is a review of a conference of the same name. Thanks to John Austin for his guiding direction in providing this paper and many of the papers summarized therein.

[19] Damasio, Antonio (2003). *Feelings of Emotion and the Self*

[20] Damasio, Antonio, (2003). *The Feeling of What Happens*, pg

[21] Ibid.

[22] Ibid, pg 182

[23] Damasio, Antonio (2003). *Feelings of Emotion and the Self*

[24] Ibid.

[25] Quote is found in Daniel Schacter's *The Seven Sins of Memory*

[26] LeDoux, Joseph, (2002). *Synaptic Self*, pg. 99

[27] LeDoux, Joseph (2003). *The Self: Clues from the Brain, pg 298*

[28] Ibid.

[29] Ibid, pg 298

[30] Schacter, Daniel, (2003) *The Seven Sins of Memory*

[31] Wegner, Daniel (2003). *The Mind's Self-Portrait*

[32] Ibid. Earlier studies by other researchers (Michotte, 1963; Einhorn & Hogarth, 1986; Kelly, 1972, 1980; McClure, 1998) focused on how people perceive causality in external events, and reached similar conclusions. A potential cause must show movement consistent with the effect, come first or at the same time as the effect, and be free from rival possible causes. If any of these conditions is missing, the sense of causation is lessened.

[33] Ibid, pgs. 217-218

[34] Gazzaniga, Michael, et al (2003) *Out of Contact, Out of Mind: The Distributed Nature of the Self*

[35] Lewis, Michael (2003) *The Emergence of Consciousness and Its Role in Human Development*

[36] for example, Huttenlocher & Higgens, 1978; Leslie, 1987; McCune-Nicolich, 1981; Piaget, 1962.

[37] Ibid, pg 387

[38] Banaji, Mahzarin & Devos, Thierry (2003). *Implicit Self and Identity*

[39] Quinn, Naomi (2002). *Cultural Selves*

[40] LeDoux, Joseph, (2002). *Synaptic Self*

[41] Markus, Hazel Rose (2003). *Models of Agency: Sociocultural Diversity in the Construction of Action*